DOCUMENT RESUME

ED 293 448 HE 021 361

AUTHOR Schwabe, Robert A.
TITLE Generating and Monitoring Faculty Allocations to Departments by Robust Regression Techniques. AIR 1987 Annual Forum Paper.
PUB DATE May 87
NOTE 24p.; Paper presented at the Annual Forum of the Association of Institutional Research (27th, Kansas City, MO, May 3-6, 1987).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Faculty; Computer Oriented Programs; *Departments; Higher Education; Institutional Research; *Regression (Statistics); *Resource Allocation; *Teacher Student Ratio
IDENTIFIERS *AIR Forum

ABSTRACT

The use of a robust statistical technique to implement a decentralized faculty allocation procedure at a medium size public university is discussed. The least median of squares (LMS), a regression technique introduced by Rousseeuw, was used instead of the usual least sum of squares, or least squares. The LMS technique was used to assist in the choice of a student-faculty ratio (SFR) for each discipline and consequently for each department. These ratios will promote a decentralized approach to the allocation of faculty on the campus. The approach requires that each of the five schools within the university accept a target full-time equivalency of students for the next school year and maintain an agreed upon SFR. Historical SFRs beginning with fall 1978 and ending with spring 1986 were computed for each discipline. Time series of at most 24 values were then analyzed by the LMS methodology as implemented in the computer program PROGRESS. While school and departmental managers may use the LMS technique to assess changes in SFRs over time by discipline, they understand that centralized monitoring will occur particularly as new allocations are made for a subsequent year. 14 references. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED293448

GENERATING AND MONITORING FACULTY ALLOCATIONS TO DEPARTMENTS BY ROBUST REGRESSION TECHNIQUES

Robert A. Schwabe
Director, Institutional Research
California State University,
San Bernardino
5500 University Parkway
San Bernardino, CA 92407
(714) 887-7291

Paper presented at the
Twenty-Seventh Annual AIR Forum
Kansas City, Missouri

Running Head: Generating and Monitoring Faculty Allocations

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
AIR
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 021 361

for Management Research, Policy Analysis, and Planning

This paper was presented at the Twenty-Seventh Annual Forum of the Association for Institutional Research held at the Vista International Hotel, Kansas City, Missouri, May 3-6, 1987. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Ann K. Dickey
Chair and Editor
Forum Publications Editorial
Advisory Committee

Teresa Karolewski
Associate Editor

ABSTRACT

Recently, this medium size public university chose to adopt a decentralized approach to the allocation of faculty. Each year the five schools are allocated a target number of full-time equivalent students with a corresponding student-faculty ratio. Subsequently, these resources are reallocated to each department. The generation and monitoring of these targets are performed centrally including the use of regression techniques applied to the appropriate time series for each discipline. These techniques include the robust method given by Peter Rousseeuw as the least median of squares regression. The use of this method is the major thrust of this paper.

INTRODUCTION

The recent change in central administration and the significant growth rate at this medium size public university has led to the decision to decentralize the allocation of faculty positions. Currently, some 7,500 students are enrolled generating 5,400 full-time equivalent students. The growth is about eight percent per year. Five schools house forty-two disciplines in thirty-five departments. Thirty-five undergraduate degree programs, five masters level degree programs and a large credential program in education is supported by more than 325 full and part-time faculty.

The proposed change is designed to move the allocation from the Office of the Vice President for Academic Affairs to the school deans. Ultimately, departmental control will come to the fore; in fact, several large departments will exercise significant influence this next academic year. The development of this decentralization will also be accompanied by a switch from a formula-based scheme to a student-faculty ratio methodology. Portends of the use of this methodology occurred in Jedamus (1974) as he discussed the difficulty of obtaining valid cross-institutional data on teaching loads usually defined as student credit hours per faculty member. He feels that this measure is at best a surrogate of an underlying reality and that longitudinal studies of these loads lead to valuable insights into an institution's operation. Dressel and Simon (1976) point to this measure when discussing budgeting

patterns and strategies. They suggest after potential student enrollments are developed, then staffing may be estimated by either: (a) setting a student credit hour per full-time equivalent faculty member productivity requirement or (b) converting course enrollments into numbers of sections.

The issues suggested by a focus on the use of student-faculty ratios often arise with respect to class size. This is of deep concern on this campus which was founded some twenty-five years ago on the premise that by far most classes should be small with a few very large classes. This premise was recently reaffirmed by our faculty senate. These concerns are addressed in Witmer (1975) as he considers opportunities for improving productivity and asserts that a most delusive goal has been increased quality of outcomes from smaller classes and lower student-faculty ratios. Feldman (1984) lays to rest the belief that class size adversely affects student evaluation of teachers and courses.

While the focus of the proposed allocation methodology is upon the use of student-faculty ratios as targets and productivity measures, the allocation process encompasses many facets of the institution's life. Budgetary issues affecting equipment, travel, non-instructional staff, physical facilities, supplies, levels of the retention, promotion and tenuring of faculty and other items of support all influence the allocation of faculty positions. Numerous decision support systems are proposed in the literature for the handling of significant numbers of these issues. For

example, Dijkman (1985) provides a comprehensive model based upon the Multiplan spreadsheet program. As to be expected, one of the normative input variables for this model is that of student-faculty ratios.

The concern of this paper is to apply a robust regression technique to assist in the choice of a student-faculty ratio for each discipline and consequently for each department. These ratios will provide a significant keystone in the development of and maintenance of a decentralized approach to the allocation of faculty on this campus.

A ROBUST REGRESSION TECHNIQUE

The robust technique used in this paper was introduced by Rousseew (1984) as a least median of squares (LMS) instead of the usual least sum of squares, or least squares (LS). This method can survive the inclusion of contamination of some fifty percent of the data, where the LS method is very sensitive to outliers such that no contamination can be tolerated. A computer program has been provided by Leroy and Rousseeuw (1985) for use on a mainframe and an IBM PC. A discussion of this program is given in the third part of this paper.

The Method of Least Squares

The method known early nineteenth century and before as the method of the least sum of squares or briefly as least squares (LS) was first published by Adrien Marie Legendre in 1805 according to Harter (1974). Harter in his massive analysis asserts concerns

for choosing a best single value to summarize information given by a number of independent observations which were in vogue with the ancient Greeks and Egyptians and with the Arabs of the Middle Ages. He traces its development and variations beginning with Galileo in 1632 through his own conclusions and recommendations in 1975. See his fascinating work in Harter (1974).

The LS method seeks to estimate an output variable y as a linear combination of k independent input variables $x_1, \ldots, x_k$ in n dimensional space. That is, one considers

$$y_i = x_{1i}\, a_1 + \ldots + x_{ki}\, a_k + e_i, \qquad i=1, \ldots, n$$

where e is assumed to be normally distributed with mean zero and standard deviation σ. The goal of the method is to determine the coefficients $a_1, \ldots, a_k$ in such a manner as to obtain the following

$$\underset{A}{\text{minimize}} \sum_{i=1}^{n} r_i^2$$

where the residuals r_i equal $y_i - x_{1i}\, a_1 - \ldots - x_{ki}\, a_k$. Numerous texts and articles discuss this approach with various levels of sophistication. See, for example, Abraham and Ledolter (1983) as well as many calculus texts and some statistics texts. The determination of the choice of the a's is made in closed form. That is, formulas exist which can be applied either manually or by computer for this determination.

It has been known for a long time that the LS methodology is susceptible to significant error with even small perturbations in the data set. The applied statistician has been aware of the

concern over outliers in most data sets. Measuring and recording errors as well as unusual occurrences reflect themselves in real data. Both aberrant y and x values cause distortions in the LS analysis. Unusually large y values affect the residuals while large x values create leverage points.

Over the years many attempts have been made to address this issue. It was not until the early 1950's that a proposal was made introducing the concept of robustness into statistical literature by G. E. P. Box according to Huber (1981). It was not until the mid-sixties that serious efforts occurred to pursue the concepts. Huber (1981) offers the following concerning a robust procedure.

(1) It should have a reasonably good efficiency.

(2) It should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly.

(3) Somewhat larger deviations from the model should not cause a catastrophe.

In addition to this book, Hampel, Ronchettl, Rousseeuw and Stahel (1986) pursue other approaches to robust techniques.

The Least Median of Squares

Rousseeuw (1984) introduced a methodology which he calls the least median of squares regression (LMS). Instead of seeking to minimize the sum of the squares of the residuals as noted above, he seeks to find the values $a_1, \ldots, a_k$ such that

$$\underset{A}{\text{minimize}} \; \underset{i}{\text{median}} \; r_i^2$$

This method meets the last two of Huber's criteria but with very low efficiency. Further, it can sustain up to 50% contamination of data. Of course, anything higher would cause concern over which half of the data is the appropriate half to use. In addition to giving a complete discussion which supports the LMS methodology from a mathematician's perspective, Rousseeuw provides several comments that are of use to the practitioner who seeks to apply it. They are: (1) The question of efficiency is not at issue for relatively small data sets since the slow convergence of the method merely suggests a longer computer run for large n. Also, he feels that the method is to be used as one of several in the process of evaluating the quality of the data set. However, he does suggest another method which produces higher efficiency. (2) While a closed form does exist for simple regression (the application of this paper), one is not known for multivariate regression. The computer implementation is designed to develop the results with the probability of 95% that "good" results are obtained (Leroy and Rousseeuw 1985). (3) The concept of a breakdown point considered to evaluate the LMS and other competing methods. This is defined as "... the smallest percentage of contaminated data that can cause the estimator to take on arbitrarily large aberrant values". For example, it is 50% for LMS and 0% for LS.

(4) In the case of simple regression, one can visualize the LMS as choosing the narrowest band about the resulting regression line which contains no less than half of the data points.

In addition to applications to several well-known data sets in statistical literature, the method has been applied to data of interest to institutional research. In the Siau, Rousseeuw and Bingen (1985) paper given at the Twenty-fifth AIR Forum in Portland, the LMS method was used to re-analyze the material which appeared in Bingen and Siau (1981). The initial article discussed the relationship between departmental activities and the need for nonacademic staff and operating funds.

ALLOCATION OF FACULTY POSITIONS

The allocation of faculty in the California State University is enrollment driven. That is, each of the nineteen campuses is allocated an annual budget based upon a given number of full-time equivalent students (FTES). An allocation model based upon recent course enrollments by the mode and level of instruction is then used to generate full-time equivalent faculty (FTEF). A campuswide student-faculty ratio (SFR) can be computed as the ratio of FTES divided by FTEF. However, for the purposes of this paper, we are interested in developing a model for generating and monitoring the allocation of faculty in the classroom only. The non-teaching use of faculty positions is handled elsewhere. Some of these uses will be staffed centrally, others at the school level and yet others at the departmental level.

An FTES is defined as 15 student credit units. For example, three students each taking a five unit course or one student taking three such courses. An FTEF supports either 12 weighted teaching units for full-time regular faculty or 15 weighted teaching units for generating part-time faculty positions. Usually four weighted teaching units are awarded a faculty member for a normal five unit course. Hence, a three course teaching load is expected for each regular faculty member. An SFR equals the ratio: FTES divided by FTEF. For most courses this is also equal to the average class size at this institution because of a unique academic plan. For other campuses in the California State University system, the average class size would be 25% larger. In fact, this will also be the case here beginning with the 1988-89 academic year. Laboratory and activity courses generate different relationships among SFR and average class size.

Current Methodology: Formulas by Class Type

Currently, linear regression is applied to three years of course data by discipline to determine an FTES allocation for a subsequent year. These allocations are modified by administrative judgement and negotiation as well as the official budget allocation. In other words, some disciplines may be expected to grow at faster rates than recent history determines and other disciplines must be maintained for various academic reasons. These allocations are then prorated according to the historical data from the previous year among the various course types within the

discipline. Appropriate formulae are then used to generate FTEF. For example, if a discipline has changed course types such as from a small class format to a large class less FTEF are generated.

Proposed Methodology: Student-Faculty Ratios

The new approach requires that each of the five schools accept a target FTES for the next school year and maintain an agreed upon SFR. These SFRs need to be consistent with the overall FTES and FTEF given to the institution as a whole. That is, the sum of the FTEFs generated by applying each discipline SFR to its FTES must be equal to the school FTEF supporting its SFR and FTES. Therefore, central allocations to each school will be made with suggested allocations for each discipline. Each department will then use appropriate discipline FTES and FTEF to maintain its course offerings. For example, some dapartments offer courses in several disciplines as is the case with the Department of Foreign Languages which provides courses in Foreign Languages, French, German and Spanish. In other departments cross-listed courses occur as well.

As in previous years, FTES allocations by discipline must be determined and an acceptable SFR used to generate FTEF. The concern of this paper is to apply the LMS technique to assist in the choice of an SFR for each discipline. To start this process, historical SFRs beginning with the Fall 1978 and ending with the Spring 1986 term were computed for each discipline. These time series of at most 24 values were then analyzed by the use of the least median of squares methodology as implemented in the computer

program PROGRESS.

In addition to the graphics output, a computed range about the LMS regression line is provided using 2.5 times the estimated scale parameter σ. This range can therefore be used to develop a sense of control when considering the SFR for a subsequent term which is not in the original data set. Of the 48 disciplines or composites thereof which were analyzed, 22 were found to have outliers, 17 generated the LS only and the remaining 9 did not have enough historical data to give meaningful results.

Figures 1 thru 4 give the graphics output for four selected disciplines. The selection was made to illustrate the several

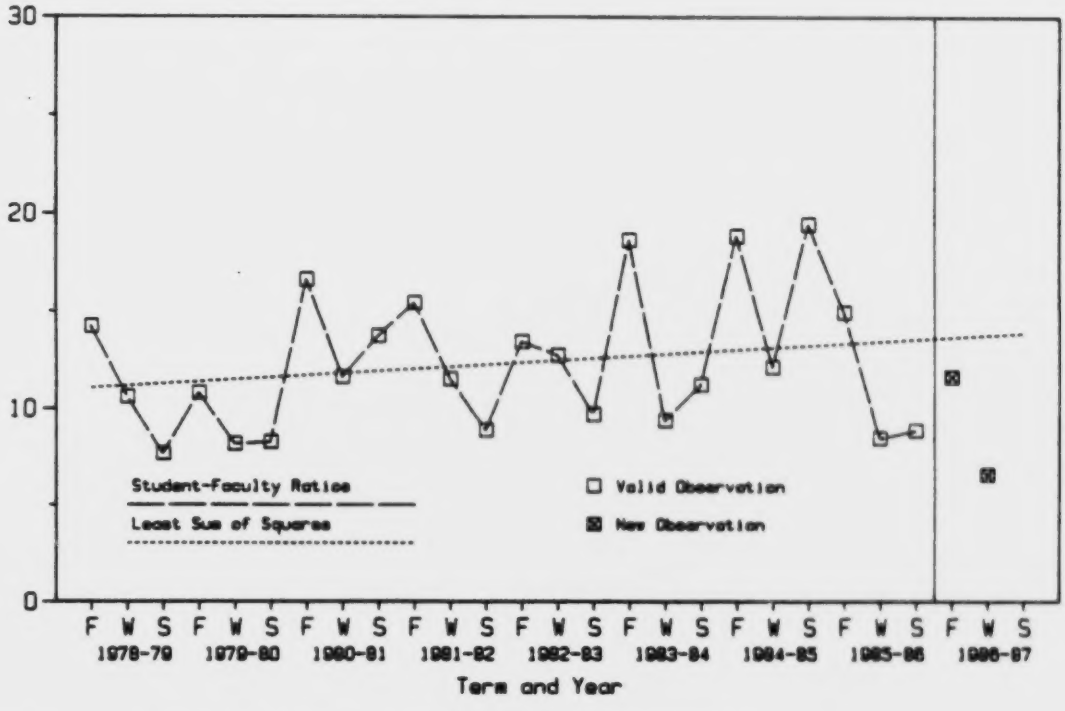

Figure 1. Music without outliers is in control.

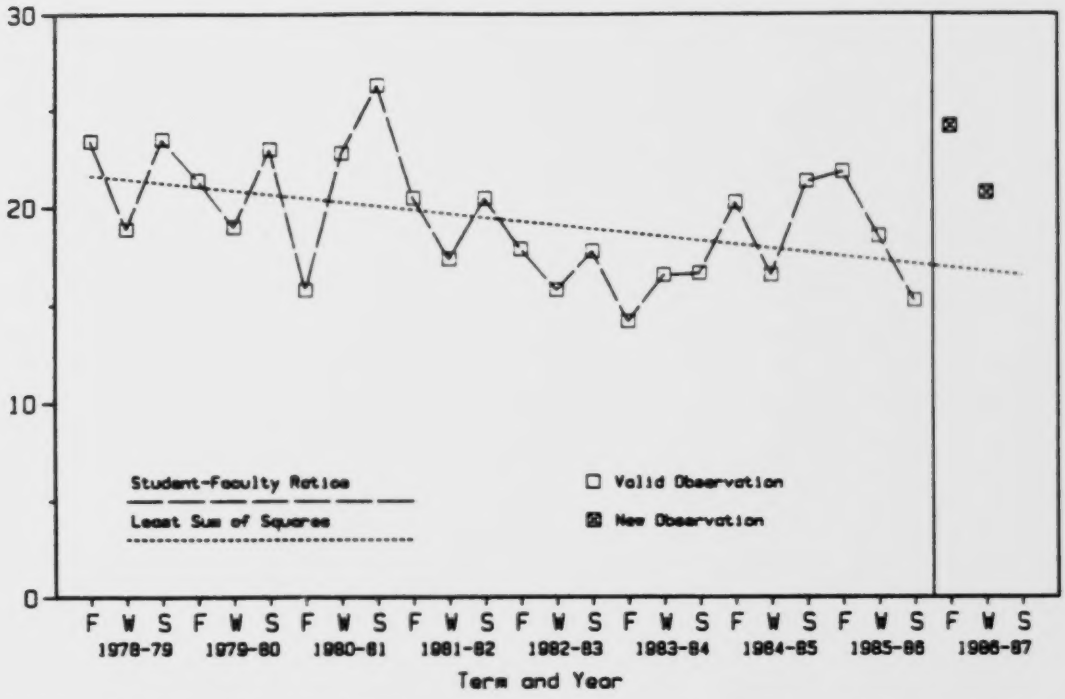
30
20
10
0
Student-Faculty Ratio
Least Sum of Squares
Valid Observation
New Observation
F W S F W S F W S F W S F W S F W S F W S F W S F W S F W S
1978-79 1979-80 1980-81 1981-82 1982-83 1983-84 1984-85 1985-86 1986-87
Term and Year

Figure 2. Interdisciplinary Study without outliers is out of control.

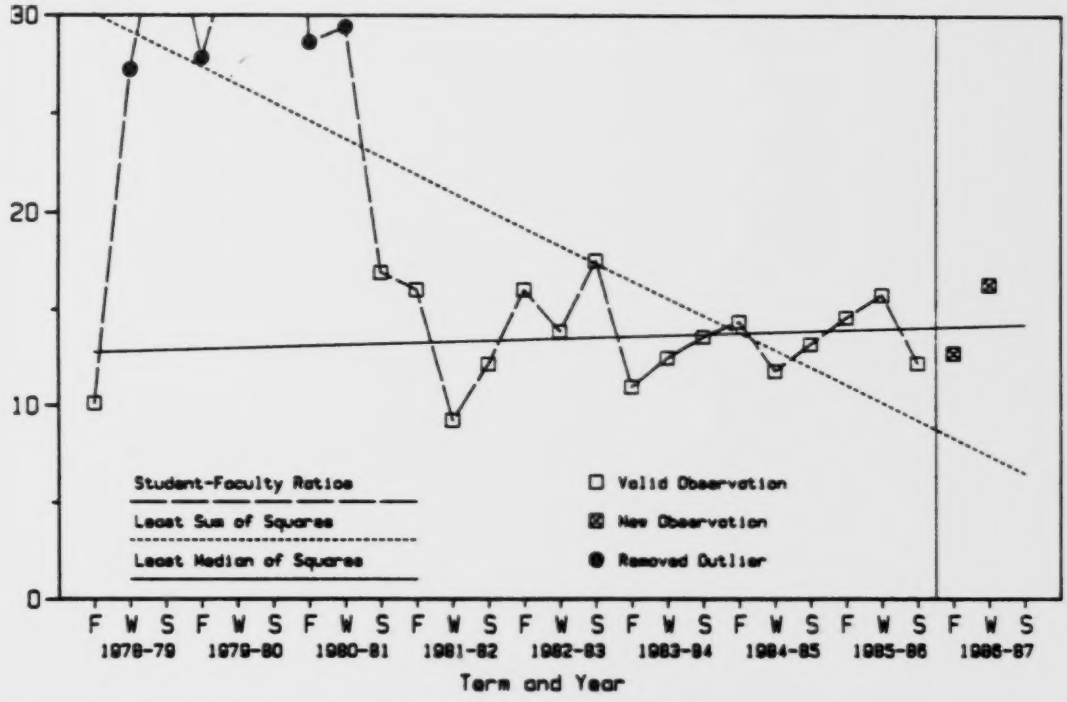
30
20
10
0
Student-Faculty Ratio
Least Sum of Squares
Least Median of Squares
Valid Observation
New Observation
Removed Outlier
F W S F W S F W S F W S F W S F W S F W S F W S F W S F W S
1978-79 1979-80 1980-81 1981-82 1982-83 1983-84 1984-85 1985-86 1986-87
Term and Year

Figure 3. Art with outliers is in control.

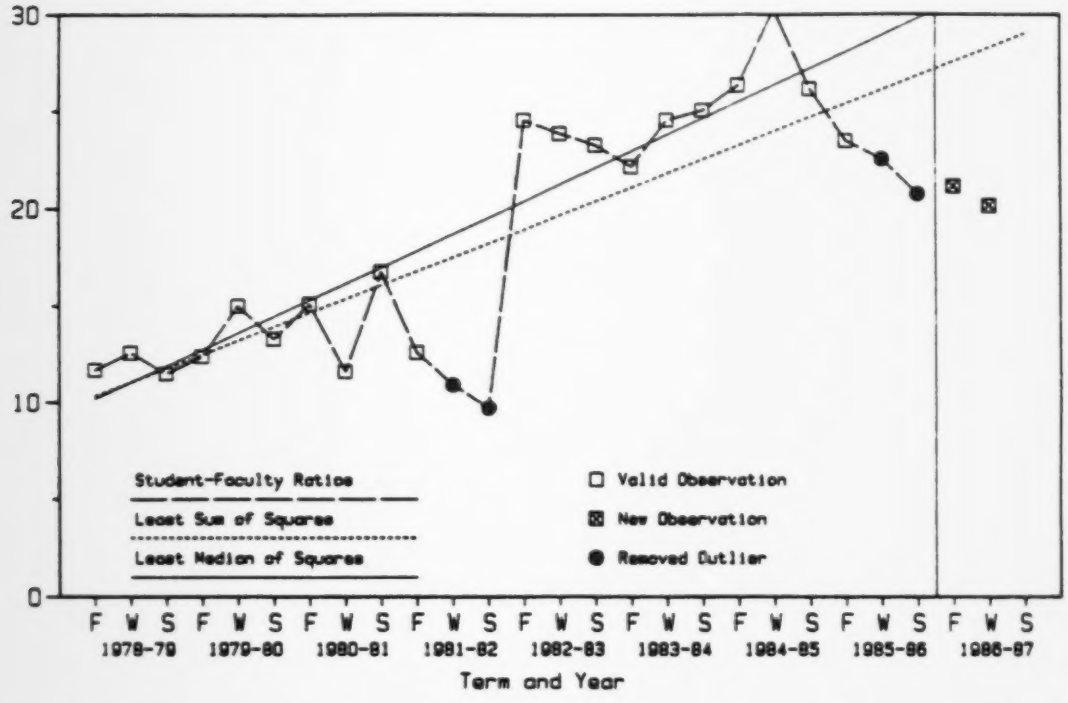
30
20
10
0
Student-Faculty Ratios
Least Sum of Squares
Least Median of Squares
Valid Observation
New Observation
Removed Outlier
F W S F W S F W S F W S F W S F W S F W S F W S F W S F W S
1978-79 1979-80 1980-81 1981-82 1982-83 1983-84 1984-85 1985-86 1986-87
Term and Year

Figure 4. Political Science with outliers is out of control.

types of output, data with and without deleted outliers and results which indicate that the Fall 1986 and Winter 1987 SFRs are in or out of control at the 98% level.

These graphs depict the time series for the past eight years. The standard least squares regression is shown as a dotted line. The robust least median of squares is shown as a solid line. All of the given SFR values are plotted in the line graph, the outliers are noted with a dot and the values chosen by the LMS algorithm are shown with a square. Two values are also plotted for the first two terms of the current academic year. They are not used to compute either the LS or the regression associated with the LMS. Table 1 summarizes these results including the control information. The

Table 1

An Analysis of Four Selected Disciplines Illustrating Those With and Without Outliers and Those In and Out of Control.

	Fall 1986	Winter 1987	Outliers	Control
Music				
SFR	11.6	6.6		
LS	13.6	13.7	No	
Residual	-2.0	-7.1		
Residual/Scale	-0.6	-2.0		Yes
Interdisciplinary Study				
SFR	24.2	20.8		
LS	17.0	16.8	No	
Residual	7.2	4.0		
Residual/Scale	2.5	1.4		No
Art				
SFR	12.6	16.2		
LMS	14.0	14.1	Yes	
Residual	-1.4	-2.1		
Residual/Scale	-0.6	-0.9		Yes
Political Science				
SFR	21.2	20.2		
LMS	30.6	31.5	Yes	
Residual	-9.4	-11.3		
Residual/Scale	-3.6	-4.3		No

SFRs for the Fall 1986 and Winter 1987 are given, the corresponding values are computed from either the LS or LMS estimate with the resulting residuals and the residual/scale. If the quotient residual/scale is strictly within the range -2.5 to 2.5, then we assert that the new SFRs are in control at the 98% level.

A COMPUTER PROGRAM

A computer program for the implementation of the least median of squares algorithm is given as PROGRESS (a Program for RObust reGRESSion) in Leroy and Rousseeuw (1985). The version presented is for an IBM PC or compatible with at least 256k RAM. The arrays will support at most 400 cases with 10 coefficients. In Leroy and Rousseeuw (1984) it is noted the program is written in FORTRAN and has passed the PFORT portability verifier; hence, it should run on any FORTRAN IV or FORTRAN 77 compiler. The author has found this to be the case and has modified it to run on a CDC Cyber 830 in a batch mode. This mode was more convenient than using the interactive mode of the IBM PC version. The SFRs were computed from standard files available on the mainframe for each of the 48 disciplines or composites that were to be analyzed. Separate files were then made to be run as input to the program PROGRESS. The output was modified to include data files necessary for input into the graphics package. The normal output was placed in a print file for further analysis. Both the input and output features of PROGRESS are discussed below.

The program proceeds interactively by asking for such items

as: (1) number of cases; (2) number of input variables; (3) title and labels for the report and variables including the dependent variable; (4) source of the input values including format - file or terminal; (5) file upon which to save input data if it was entered by terminal; (6) print option for either (a) basic results, (b) includes (a) with observed and estimated values with the residuals and weights, or (c) includes (b) with raw and standardized data; (7) residual plots as (a) none, (b) standardized residuals versus estimated value of Y, (c) standardized residuals versus index of the observation (d) both (b) and (c); (8) quick or extensive algorithm; and (9) treatment of missing values.

The program follows the pattern of computing the LS after standardizing the data by using medians and a scale estimate for σ. The LMS algorithm is then applied to determine either outliers or leverage points. Those that are selected are given a weight of 0.0 as opposed to 1.0 for the remaining data points. Lastly, LS is rerun on the data that was chosen as valid by LMS. Figures 5 and 6 are the corresponding residual plots for the first two steps using the Political Science data discussed above since the plot associated with the reweighted LS run is very similar in appearance to the LMS plot in this case.

The balance of the output includes other items of interest such as: (1) data lists at each step; (2) Spearman and Pearson correlation coefficients; (3) the variance and covariance matrix; (4) coefficient of determination (R squared); and (5) medians,

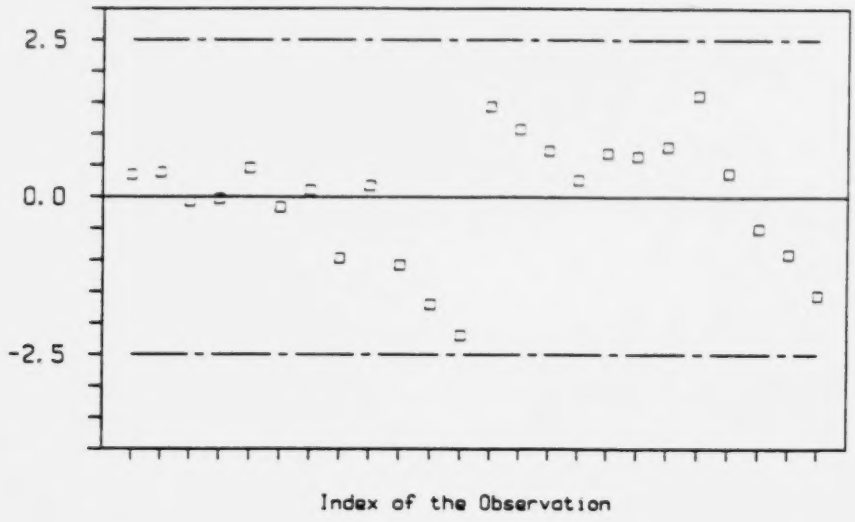

Figure 5. Standardized residuals versus the index of the observation based upon the initial run of the least squares algorithm showing no outliers.

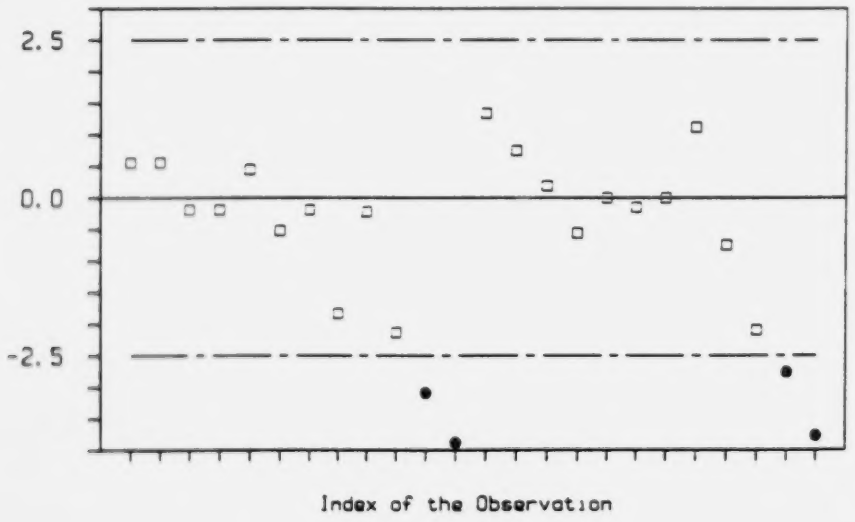

Figure 6. Standardized residuals versus the index of the observation based upon the least median squares algorithm indicating outliers.

dispersions, sum of squares with scale estimates of σ, t-values and standard errors.

CONCLUSION

The robust statistical technique introduced as the least median of squares by Rousseeuw (1984) has been found to be useful for this medium size university as it implements a decentralized faculty allocation procedure. The process of helping school deans and department chairs to understand some of the ramifications of the meaning and use of a student-faculty ratio as a departmental management tool is aided by the creation of a simple regression analysis of the changes in these ratios over time by discipline. This analysis is enhanced by the use of this technique and its computerized implementation since it is as Gardner (1981) stated when discussing a related technique "... easily understood, simple to apply, and, in many situations, reasonably accurate". It helps isolate aberrant data which must be explained or otherwise handled as well as supports an approach which gives control levels for future changes. School and departmental managers may at their discretion moderate these changes as it suits their needs with the understanding that centralized monitoring will occur particularly as new allocations are made for a subsequent year.

The least median of squares technique will also be applied to the other portions of the allocation procedure which generates the necessary full-time equivalent student target allocation to each discipline. This target allocation and an updated student-faculty

ratio will provide for the number of faculty positions devoted to classroom activities. Other uses of faculty positions will be determined at the school and departmental level.

The advent of the micro-computer with its plethora of software providing decision support systems has enhanced the move toward decentralized control of faculty and other resources. It is also important to have available new and improved analytic tools which are based in the recently developed area of robust statistical techniques.

References

Bingen, F., and Siau, C. (1981). An analysis of the needs for nonacademic staff and operating funds in a university. Research in Higher Education, 15 (1): 31-48.

Dressel, P., and L. A. K. Simon. (1976). Allocating resources among departments. New Directions in Institutional Research, 11.

Dijkman, F. (1985). An allocation model for teaching and nonteaching staff in a decentralized institution. Research in Higher Education, 22 (1): 3-18.

Feldman, J. A. (1984). Class size and college student's evaluation of teachers and courses: a closer look. Research in Higher Education, 21 (1): 45-116.

Gardner, D. (1981). Weight factor selection in double exponential smoothing enrollment forecasts. Research in Higher Education, 14 (1): 49-56.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). Robust Statistics. New York: Wiley.

Harter, H. L. (1974). The method of least squares and some alternatives. International Statistics Review, 42: 147-174, 235-264, 43: 1-44, 125-190, 269-278.

Huber, P. J. (1981). Robust Statistics. New York: Wiley.

Jedamus, P. (1974). Teaching loads over time. In J. I. Doi (Ed.), Assessing faculty effort. New Directions in Institutional Research, 2: 17-34.

Leroy, A., and Rousseeuw, P. (1984). A multiple regression technique for detecting outliers. Research Report No. 84-33, Department of Mathematics and Infromatics, University of Brussels.

Leroy, A., and Rousseeuw, P. (1985). PROGRESS: a program for robust regression analysis on the IBM PC. Research Report No. 85-14, Department of Mathematics and Infromatics, University of Brussels.

Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American Statistical Association, 79, (388): 871-880.

Siau, C., Rousseeuw, P., and Bingen, F. (1985). Applying robust regression techniques to institutional data. Paper presented at the 25th Forum of the Association for Institutional Research, Portland.

Witmer, D. R. (1975). Opportunities for improving productivity. In R. A. Wallhaus (Ed.), Measuring and increasing academic productivity. New Directions in Institutional Research, 8: 117-129.